Adaptively Regularized Phase Alignment Precoding for Multiuser Multiantenna Downlink

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

Adaptively Regularized Phase Alignment Precoding for Multiuser Multiantenna Downlink

S. Morteza Razavi and Tharmalingam Ratnarajah, *Senior Member, IEEE*

***Abstract*—In multiantenna downlink, linear precoders like channel inversion (CI), regularized CI (RCI), and phase alignment (PA) are more desirable than their nonlinear counterparts due to their reduced complexity. However, to exploit the full benefits of multiuser downlink, the availability of perfect channel state information (CSI) at the base station (BS) is mandatory. Since such an assumption is not realistic in practice, subject to CSI mismatch, the downlink precoders may undergo a severe degradation in performance. Therefore, an adaptive linear precoder is required to achieve better performance under the availability of imperfect CSI at the BS. In this work, we propose such a linear precoder by judiciously picking up an optimized regularization parameter based on the knowledge of the channel estimation error variance in advance. To do so, we first introduce a generalized CSI mismatch model where the variance of the channel estimation error is a function of signal-to-noise ratio (SNR) and is thus able to accommodate a variety of distinct scenarios like reciprocal channels and CSI feedback. It is shown that the proposed precoding scheme, dubbed adaptively regularized PA (RPA), is capable of achieving lower bit error rates (BER) compared to standard linear precoders under both perfect and imperfect CSI.**

***Keywords*—Imperfect CSI, linear precoding, multiuser multiantenna downlink, regularized phase alignment.**

I. Introduction

DEPLOYING multiple antennas at the base station (BS) can significantly improve the achievable throughput in broadcast channels so that many users can be simultaneously serviced. In this case, dirty paper coding (DPC) technique is capable of achieving the downlink capacity [1]. Nevertheless, due to its high complexity, some less complex *nonlinear* precoders like vector perturbation [2] and Tomlinson-Harashima [3] are also of particular interest.

Although achieving less throughput, *linear* precoders are more practical because of the reduced complexity compared to their *nonlinear* counterparts. The least complex and the most well-known technique, named channel inversion (CI) [4], is a linear precoding technique which yields reasonable performance in downlink. Nonetheless, it has been shown that CI precoding, while generally suboptimal, can achieve the same asymptotic sum capacity as DPC, while the number of users goes to infinity [5]. However, when the number of antennas at the BS is equal to the total number of single-antenna users and both are finite, it has been shown that with proportionally increasing the number of users and the number of antennas at the BS, the bit error rate (BER) experienced by each user due to deploying CI precoding deteriorates. This inferior performance of CI is related to the erratic behavior of the largest eigenvalue of the inverse of the covariance matrix of the channel. One approach to alleviate this misbehavior of CI precoding is to exploit the concept of regularization by adding a multiple of the identity matrix to the covariance matrix of the channel before inverting. This precoding, which has been dubbed regularized CI (RCI) [4], improves the performance of CI, such that now with proportionally increasing the number of users and the number of antennas at the BS, the BER experienced by each user remains fixed at low-to-intermediate signal-to-noise ratios (SNRs), and slightly improves at high SNRs.

In line with the aim of linear precoders to achieve reduced complexity compared to nonlinear precoders, [6] proposed a linear precoding technique based on correlation rotation (which is hereafter called phase alignment (PA)) where instead of removing the harmful symbol-to-symbol interference, it rotates the phases of the transmitted symbols such that the destructive interference becomes constructive, and eventually leading to more output SNRs for a fixed transmit power at the BS. Consequently, PA can significantly reduce the BER experienced by each user compared to CI and RCI in broadcast channels. The superior performance of PA precoding compared to standard precoders has been also investigated in [7] for cognitive radio networks and in [8] for large-scale multiantenna transmitters.

However, this superior performance of PA to CI and RCI precoders can be gleaned only when perfect CSI is available at the BS. Form the practical point of view, on the other hand, having access to perfect CSI is not realistic, and in this case, the performance of CI, RCI and PA may become adversely affected. Therefore, designing an adaptive linear precoder which outperforms standard linear precoders under imperfect CSI is of particular interest.

In this work we aim to do so by first introducing a generalized CSI mismatch model where the variance of the channel estimation error depends on the SNR. This imperfect CSI model covers a variety of practical scenarios like reciprocal channels and CSI feedback such that in the case of the latter, there is a strong connection between the error variance of the channel estimation and the size of the codebook (or the number of fedback bits).

RCI precoding improves the performance of CI by only

This work was supported by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number: HíATUS-265578.

S. M. Razavi was with the Institute for Digital Communications, The King's Buildings, The University of Edinburgh, UK, email: Morteza.Razavi@ed-alumni.net .

T. Ratnarajah is with the Institute for Digital Communications, The King's Buildings, The University of Edinburgh, UK, email: T.Ratnarajah@ed.ac.uk .

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

adding a multiple of the identity matrix to the covariance matrix of the channel before inverting (i.e., regularization). For the case of very large number of transmit and receive antennas, different regularization parameters could be derived under different scenarios for RCI precoding (see e.g., [9]–[11]). Similarly, in this paper, we propose an enhanced PA precoding, dubbed adaptively regularized PA (RPA), by deriving an optimized regularization parameter which is a function of the channel estimation error variance. The proposed approach is adaptive in a sense that the sought regularization parameter is based on the knowledge of the channel estimation error variance which is possible to be known in advance due to the channel dynamics and channel estimation schemes. Considering the fact that in this paper we address the *linear* downlink precoders, we can summarize the contributions of this paper as follows:

- Although the concept of RPA precoding has been primarily introduced in [12], the *adaptive* RPA precoder proposed in this work holds two main advantages: First, the *standard* RPA precoding proposed in [12] is only applicable to the case of *equal* number of transmit and receive antennas. However, here we generalize the concept of RPA precoding to the case of *unequal* number of transmit and receive antennas. Second, the *standard* RPA precoder has been only analyzed in the presence of perfect CSI at the BS. The *adaptive* RPA precoding in this work, on the other hand, considers the the characteristics of the CSI imperfections to design a more efficient precoder which eventually leads to the much better performance under CSI mismatch. It is worthy to note that although this generalization of the RPA precoding is not trivial, the proposed approach does not introduce any extra computational complexity compared to the one in [12].
- It is shown that the *adaptive* RPA precoder proposed in this work is able to outperform the other linear precoders, i.e., CI, RCI, and PA precoding, under both perfect and imperfect CSI including reciprocal channels and CSI feedback. More specifically, the larger the M and N, the better performance can be gleaned by using adaptive RPA precoding. It is also worthy to note that within the practical SNR regime, adaptive RPA outperforms RCI even for the case of unequal number of transmit and receive antennas.

A. Paper Organization

In Section II, we overview the system model of standard downlink precoding followed by a generalized model for imperfect CSI in Section III. In Section IV, we propose an adaptive RPA precoding which is able to outperform standard linear precoders. In Section V and by using numerical simulations, we demonstrate the superior performance of the proposed scheme. Finally, Section VI contains a summary of the presented materials within the paper.

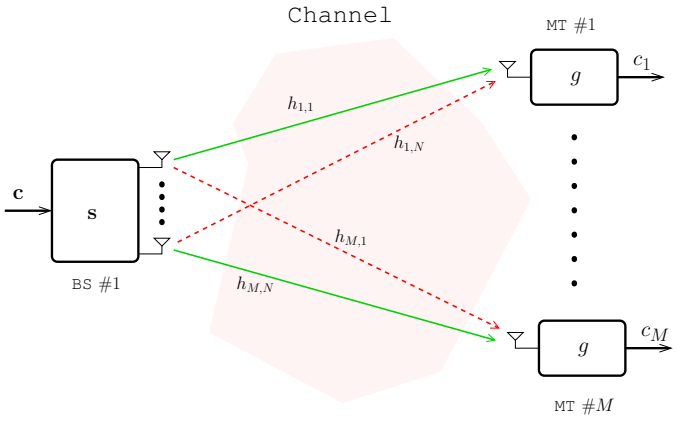

Fig. 1. Single-cell broadcast channel where dash red arrows represent intra-cell interference while solid green arrows denote desired links. $h_{k,j}$ is the time-variant channel response between the jth transmit antenna of the BS and the kth MT.

B. Notations

Throughout the paper, a is a scalar, $\mathbf{a}$ is a vector, and $\mathbf{A}$ is a matrix. The superscripts $(\cdot)^{\mathsf{H}}$ and $(\cdot)^{\top}$ represent the Hermitian transpose and the non-conjugated transpose, respectively. $\mathbb{E}\{\cdot\}$ and $\mathrm{Tr}[\cdot]$ are the expectation and trace operators, respectively. While $\|\cdot\|_2$ denotes the vector 2-norm, $\odot$ designates the Hadamard (element-wise) matrix product, and $|\cdot|$ represents the element-wise absolute value.

II. System Model

We consider a multiuser downlink scenario where an N-antenna transmitter communicates with mobile terminals (MTs) with M receive antennas in total, as depicted in Fig. 1. More specifically, we assume that the total number of receive antennas is equal to or less than the total number of transmit antennas, i.e., $M \leq N$. Since no signal processing treatment is going to be considered at each MT, the system configuration is irrespective to whether the receive antennas cooperate or not, therefore the total number of receive antennas can belong to one user or be shared by several users; however, as purely transmitter-based precoders are most useful with single antenna receivers, we consider single-antenna MTs in the following, which is also consistent with the references of this paper.

Without loss of generality, we assume that all single-antenna users are homogeneous and experience independent fading. The received signals of all users can be collectively shown by

$$\mathbf{y} = \sqrt{P}\mathbf{H}\mathbf{s} + \mathbf{z} \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{M\times 1}$, P is the transmit power, and $\mathbf{H} \in \mathbb{C}^{M\times N}$ denotes the channel from N-antenna transmitter to M single-antenna users such that the magnitude of channel coefficients is bounded away from zero and infinity. We also consider block fading model where channel coefficients are static for the duration of a transmission but may change between successive transmission. We further assume that elements of $\mathbf{H}$ can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

modeled by independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance, i.e., $\mathrm{vec}\,(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \mathbf{I})$, $\mathbf{s} \in \mathbb{C}^{N\times 1}$ is the transmitted signal from the BS, and $\mathbf{z} \in \mathbb{C}^{M\times 1}$ is the circularly symmetric additive white Gaussian noise with zero mean and variance σ^2, i.e., $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$. We further assume that the transmitted signal $\mathbf{s}$ in (1) can be expressed as $\mathbf{s} = g\mathbf{\Psi}\mathbf{c}$. Similar to [4], [12], [13], we consider g as the scaling factor that ensures transmit power constraint i.e., $\mathbb{E}\left\{\|\mathbf{s}\|_2^2\right\} = 1$. $\mathbf{\Psi}$ is the precoding matrix and $\mathbf{c}$ represents the vector containing the symbols chosen from a desired PSK constellation and since we assume i.i.d. input signaling, we have $\mathbb{E}\{\mathbf{C}\} = \mathbf{I}$, where $\mathbf{C} = \mathbf{c}\mathbf{c}^{\mathsf{H}}$ is the covariance matrix of the transmitted data stream. We also define the nominal SNR as $\gamma = P/\sigma^2$. Note that although the concept of regularization is most beneficial for the case of equal number of transmit and receive antennas [4], [12], without loss of generality, we assume $M \leq N$.

III. Imperfect CSI Model

Similar to [14], we model the imperfect CSI at the transmit side as

$$\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{E} \tag{2}$$

where the actual channel matrix $\mathbf{H}$ is thought to be independent of channel measurement error $\mathbf{E}$. We further consider $\mathbf{E}$ as a Gaussian matrix consisting of i.i.d. elements with mean zero and variance τ such that [14]

$$\mathrm{vec}\,(\mathbf{E}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \tau\mathbf{I}) \;\text{ with } \tau \triangleq \beta\gamma^{-\alpha}, \quad \beta > 0, \quad \alpha \geq 0 \tag{3}$$

In this case, the error variance can depend on SNR $(\alpha \neq 0)$ or be independent of that $(\alpha = 0)$. In particular, perfect CSI is regained by setting $\tau = 0$. Notice the variance model in (3) is versatile since it is potentially able to accommodate a variety of distinct scenarios. More specifically, τ can be interpreted as a parameter that captures the quality of the channel estimation which is possible to be known *a priori*, depending on the channel dynamics and channel estimation schemes (see e.g., [15] and references therein). Two representative cases of this error variance model can be described as follows [14]:

- *Reciprocal Channels:* This case represents the reciprocal systems like time division duplex where uplink and downlink channels are identical. The downlink channel can thus be estimated through pilots sent over the uplink channel and the channel measurement error $\mathbf{E}$ depends on the noise level at the BS as well as the pilot power. If the pilot power proportionally increases with P, the channel estimation error scales inversely with increasing γ. This case is modeled by setting $\alpha = 1$.
- *CSI Feedback:* Here, the channel matrix can be estimated by pilot transmissions in the downlink. Then, a quantized version of this channel estimate is sent back to the BS through a dedicated feedback link. This way, the imperfect CSI will be mostly dominated by the errors caused through quantization and feedback delay, which can eventually result in outdated CSI at the BS if the channel coherence time is smaller than the feedback delay. Since channel coherence time and the resolution of quantizer do not depend on γ, the channel estimation error variance τ becomes independent of γ as well. This case is captured by setting $\alpha = 0$. For instance, if each user has a codebook of size 2^B designed based on random vector quantization (RVQ) [16]–[18], and feeds back the index of a quantized complex channel vector of length N to the BS by using B bits, it has been shown that the error variance τ is directly related to B and N as $\tau = 2^{-B/N}$ [19].

To further facilitate the performance analysis of the linear precoders under the CSI mismatch model in (2), it is more appropriate to have the statistical properties of $\mathbf{H}$ conditioned on $\widehat{\mathbf{H}}$ by using following lemma [20]:

Lemma 1: Conditioned on $\widehat{\mathbf{H}}$, $\mathbf{H}$ has a Gaussian distribution with mean $\widehat{\mathbf{H}}/\,(1+\tau)$ and statistically independent elements of variance $\tau/\,(1+\tau)$, i.e.,

$$\mathbf{H} = \frac{1}{1+\tau}\widehat{\mathbf{H}} + \breve{\mathbf{H}} \tag{4}$$

where the auxiliary random matrix $\mathrm{vec}\left(\breve{\mathbf{H}}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{\tau}{1+\tau}\mathbf{I}\right)$ is statistically independent of $\widehat{\mathbf{H}}$.

IV. Adaptively Regularized Phase Alignment Precoding

Since in practice, acquiring perfect CSI is not pragmatic and only partial CSI may be accessible, in this section, we propose an adaptive RPA precoding by deriving an optimized regularization parameter which is chosen based on the knowledge of the channel estimation error variance in advance, as discussed in Section III. Therefore, in the following, we assume that instead of perfect CSI, i.e., $\mathbf{H}$, only imperfect CSI $\widehat{\mathbf{H}}$ is available at the BS, where $\widehat{\mathbf{H}}$ and $\mathbf{H}$ are related to each other through equations (2) and (4).

As mentioned earlier and similar to RCI precoding which consists of an appropriate regularization parameter to further improve the performance of CI precoding, the proposed RPA precoding seeks a regularization parameter to enhance the performance of PA precoding. Thus to further proceed, PA precoding should first be considered.

A. PA Precoding under Perfect CSI

Under the availability of perfect CSI at the BS and by using PA precoding, the transmitted signal from the BS can be described as [6]

$$\mathbf{s}_{\mathrm{PA}} = g_{\mathrm{PA}}\mathbf{\Psi}_{\mathrm{PA}}\mathbf{c} \tag{5}$$

where

$$\mathbf{\Psi}_{\mathrm{PA}} = \mathbf{H}^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{R}_{\theta} \tag{6}$$

is the precoding matrix and $\mathbf{R} = \mathbf{H}\mathbf{H}^{\mathsf{H}}$ is the covariance matrix of the channel. Plus, we have [12]

$$\mathbf{R}_{\theta} = |\mathbf{R}| \odot \mathbf{C} \tag{7}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

TABLE I. STATISTICAL PROPERTIES OF $\left|\rho_{\ell,x}\right|$, WHERE $\rho_{\ell,x}$ IS THE (ℓ,x)-TH ELEMENT OF $\boldsymbol{\mathcal{R}}=\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathsf{H}}$ AND $\mathrm{vec}\left(\boldsymbol{\mathcal{H}}\right)\sim\mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\sigma_h^2\mathbf{I}\right)$ (APPENDIX).

	Type of Random Variable	$\mathbb{E}\left\{\left\|\rho_{\ell,x}\right\|\right\}$	$\mathbb{E}\left\{\left\|\rho_{\ell,x}\right\|^2\right\}$
$\ell\neq x$	Rayleigh	$\sigma_h^2\sqrt{N\pi}/2$	$\sigma_h^4 N$
$\ell=x$	Chi-square	$\sigma_h^2 N$	$\sigma_h^4 N\left(N+1\right)$

and [6]

$$g_{\mathtt{PA}}=\frac{1}{\sqrt{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]}} \tag{8}$$

is the scaling factor.

B. RPA Precoding under Imperfect CSI

In the presence of CSI mismatch, the transmitted signal from the BS in (1) can be shown as

$$\widehat{\mathbf{s}}_{\mathtt{RPA}}=\widehat{g}_{\mathtt{RPA}}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{R}}_\theta\mathbf{c} \tag{9}$$

such that

$$\widehat{\mathbf{R}}_\theta=\left|\widehat{\mathbf{R}}\right|\odot\mathbf{C}\quad,\quad\widehat{\mathbf{R}}=\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} \tag{10}$$

We further define

$$\widehat{\boldsymbol{\Psi}}_{\mathtt{RPA}}=\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{R}}_\theta \tag{11}$$

and

$$\widehat{\mathbf{c}}=\widehat{\mathbf{R}}_\theta\mathbf{c} \tag{12}$$

This way, the scaling factor can be represented as

$$\widehat{g}_{\mathtt{RPA}}=\frac{1}{\sqrt{\mathrm{Tr}\left[\widehat{\boldsymbol{\Psi}}_{\mathtt{RPA}}^{\mathsf{H}}\widehat{\boldsymbol{\Psi}}_{\mathtt{RPA}}\right]}} \tag{13}$$

By considering (1) and (9), the sought precoder minimizes the mean square error between the transmitted data stream and the received signal plus noise. Therefore, under imperfect CSI, the desired precoder can be found by using the following optimization criterion:

$$\min_{\ddot{\boldsymbol{\Psi}}}\mathbb{E}\left\{\left\|\sqrt{P}\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}+\ddot{f}\mathbf{z}-\sqrt{P}\widehat{\mathbf{c}}\right\|_2^2\right\} \tag{14}$$

and

$$\ddot{f}=\frac{1+\tau}{\widehat{g}_{\mathtt{RPA}}}=\left(1+\tau\right)\sqrt{\mathrm{Tr}\left[\widehat{\boldsymbol{\Psi}}_{\mathtt{RPA}}^{\mathsf{H}}\widehat{\boldsymbol{\Psi}}_{\mathtt{RPA}}\right]} \tag{15}$$

Note that the inclusion of $\ddot{f}$ in (14) is due to the fact that in all precoding schemes, the power of noise is affected by the precoding matrix, and consequently, this effect can be reflected through a multiplicative factor like $\ddot{f}$. This can be perceived with respect to the fact that at transmit side, the transmitted signals get scaled by $\widehat{g}_{\mathtt{RPA}}$ to meet the power constraints; consequently at receive side, to have an unbiased detection, the received signals should be scaled back by $\left(1+\tau\right)/\widehat{g}_{\mathtt{RPA}}$ which further appears as a multiplicative factor for the noise vector. More explanation on why $\widehat{g}_{\mathtt{RPA}}$ needs to be scaled up by $\left(1+\tau\right)$ is available in [14].

To further proceed, we consider the following two lemmas:

Lemma 2: $\mathbb{E}\left\{\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\right\}=\varrho\,\mathbf{I}$ where

$$\varrho=N\left(1+\tau\right)^2\left[\left(N+1\right)+\left(M-1\right)\left(1+\sqrt{N\pi}+\left(M-2\right)\frac{\pi}{4}\right)\right] \tag{16}$$

Proof: Note that (12) can be rewritten as

$$\widehat{\mathbf{c}}=\left(\begin{array}{c}c_1\sum_{x=1}^{M}\left|\widehat{\rho}_{1,x}\right|,\ldots,c_M\sum_{x=1}^{M}\left|\widehat{\rho}_{M,x}\right|\end{array}\right)^{\top} \tag{17}$$

Also by considering (2), we have $\mathrm{vec}\left(\widehat{\mathbf{H}}\right)\sim\mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\left(1+\tau\right)\mathbf{I}\right)$. Therefore, the statistical properties of $\left|\widehat{\rho}_{\ell,x}\right|$, which is the (ℓ,x)-th element of $\widehat{\mathbf{R}}$ in (10), can be obtained by setting $\sigma_h^2=1+\tau$ in Table I. Therefore, after some straightforward manipulations and by considering the fact that $\mathbb{E}\left\{\mathbf{c}\mathbf{c}^{\mathsf{H}}\right\}=\mathbf{I}$, the claim follows. ■

Lemma 3: $\mathbb{E}\left\{\widehat{\mathbf{R}}_\theta^2\right\}=\kappa\,\mathbf{I}$ where

$$\kappa=N\left(1+\tau\right)^2\left(M+N\right) \tag{18}$$

Proof: Based on (10), we have

$$\mathbb{E}\left\{\widehat{\mathbf{R}}_\theta^2\right\}=\mathrm{diag}\left(\mathbb{E}\left\{\left[\widehat{\mathbf{R}}^2\right]_{1,1}\right\},\ldots,\mathbb{E}\left\{\left[\widehat{\mathbf{R}}^2\right]_{M,M}\right\}\right) \tag{19}$$

Considering the fact that $\widehat{\mathbf{R}}$ is a Hermitian matrix, we have

$$\left[\widehat{\mathbf{R}}^2\right]_{\ell,\ell}=\sum_{i=1}^{M}\left|\widehat{\rho}_{\ell,i}\right|^2 \tag{20}$$

Thus, based on Table I and since $\sigma_h^2=1+\tau$, we have $\mathbb{E}\left\{\left[\widehat{\mathbf{R}}^2\right]_{\ell,\ell}\right\}=N\left(1+\tau\right)^2\left(M+N\right)$. Therefore, the claim follows. ■

Due to Lemma 2 and Lemma 3, the objective function in (14) can be shown as

$$\begin{aligned}\ddot{F}&=\mathbb{E}\left\{\mathrm{Tr}\left[\left(\sqrt{P}\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}+\ddot{f}\mathbf{z}-\sqrt{P}\widehat{\mathbf{c}}\right)\right.\right.\\&\times\left.\left.\left(\sqrt{P}\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}+\ddot{f}\mathbf{z}-\sqrt{P}\widehat{\mathbf{c}}\right)^{\mathsf{H}}\right]\right\}\\&=\mathbb{E}\left\{\mathrm{Tr}\left[P\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}+\ddot{f}^2\mathbf{z}\mathbf{z}^{\mathsf{H}}-P\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\right.\right.\\&-P\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}+P\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}+\sqrt{P}\ddot{f}\mathbf{H}\ddot{\boldsymbol{\Psi}}\widehat{\mathbf{c}}\mathbf{z}^{\mathsf{H}}\\&+\left.\left.\sqrt{P}\ddot{f}\mathbf{z}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}-\sqrt{P}\ddot{f}\widehat{\mathbf{c}}\mathbf{z}^{\mathsf{H}}-\sqrt{P}\ddot{f}\mathbf{z}\widehat{\mathbf{c}}^{\mathsf{H}}\right]\right\}\\&\overset{①}{=}P\varrho\,\mathrm{Tr}\left[\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\ddot{\boldsymbol{\Psi}}\right]+M\sigma^2\left(1+\tau\right)^2\kappa\,\mathrm{Tr}\left[\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\ddot{\boldsymbol{\Psi}}\right]\\&-P\varrho\,\mathrm{Tr}\left[\mathbf{H}\ddot{\boldsymbol{\Psi}}\right]-P\varrho\,\mathrm{Tr}\left[\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right]+PM\varrho\end{aligned} \tag{21}$$

where ① follows the fact that the noise vector $\mathbf{z}$ is considered to be independent of data stream $\mathbf{c}$ and the perfect CSI $\mathbf{H}$. To obtain the sought precoder, we can differentiate $\ddot{F}$ with respect to $\ddot{\boldsymbol{\Psi}}$ by first considering the following assumptions [21]:

1) $\ddot{\boldsymbol{\Psi}}$ and $\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}$ are treated as mutually independent variables such that $\dfrac{\partial\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}}{\partial\ddot{\boldsymbol{\Psi}}}=\dfrac{\partial\ddot{\boldsymbol{\Psi}}}{\partial\ddot{\boldsymbol{\Psi}}^{\mathsf{H}}}=\mathbf{0}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

2) $\dfrac{\partial \operatorname{Tr}\left[\mathbf{A}\ddot{\mathbf{\Psi}}\right]}{\partial \ddot{\mathbf{\Psi}}} = \dfrac{\partial \operatorname{Tr}\left[\ddot{\mathbf{\Psi}}\mathbf{A}\right]}{\partial \ddot{\mathbf{\Psi}}} = \mathbf{A}.$

Following the preceding assumptions, the differentiation of $\ddot{F}$ with respect to $\ddot{\mathbf{\Psi}}$ gives

$$\begin{aligned}\frac{\partial \ddot{F}}{\partial \ddot{\mathbf{\Psi}}} &= P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H} + M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} - P\varrho\mathbf{H} \\ &\overset{②}{=} P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right)^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right) \\ &+ M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} - P\varrho\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right) \\ &= P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\left[\frac{\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}}{(1+\tau)^2}+\breve{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}+\frac{\widehat{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}+\breve{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}}{1+\tau}\right] \\ &+ M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} - \frac{P\varrho}{1+\tau}\widehat{\mathbf{H}} - P\varrho\breve{\mathbf{H}} \qquad (22)\end{aligned}$$

where ② follows from (4). The adaptive precoder can then be found by setting $\partial\ddot{F}/\partial\ddot{\mathbf{\Psi}}$ equal to zero and taking the expectation over the auxiliary random matrix $\breve{\mathbf{H}}$. First, note that we have

$$\mathbb{E}\left\{\breve{\mathbf{H}}\right\} \overset{③}{=} \mathbf{0} \qquad (23a)$$

$$\mathbb{E}_{\breve{\mathbf{H}}}\left\{\widehat{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}\right\} \overset{④}{=} \mathbb{E}_{\breve{\mathbf{H}}}\left\{\breve{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}\right\} = \mathbf{0} \qquad (23b)$$

$$\mathbb{E}\left\{\breve{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}\right\} \overset{⑤}{=} \frac{M\tau}{1+\tau}\mathbf{I} \qquad (23c)$$

where ③ and ④ follow Lemma 1, and ⑤ is due to the subsequent lemma:

Lemma 4: If $\mathbf{A} \in \mathbb{C}^{M\times N}$ represents a Gaussian matrix with i.i.d. elements of mean zero and variance a, then $\mathbb{E}\left\{\mathbf{A}\mathbf{A}^{\mathsf{H}}\right\} = aN\,\mathbf{I}$.

Proof: Since $\mathbf{A}$ is a Gaussian matrix, we have $\operatorname{vec}(\mathbf{A}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, a\mathbf{I})$. In other words, if $\mathbf{a}$ represents an arbitrary column of $\mathbf{A}$, then $\mathbb{E}\left\{\mathbf{a}\mathbf{a}^{\mathsf{H}}\right\} = a\,\mathbf{I}$ [22]. However, since $\mathbf{A}$ has N independent columns, the claim follows. ■

Therefore, we can represent the precoding matrix as

$$\begin{aligned}\mathbb{E}_{\breve{\mathbf{H}}}\left\{\frac{\partial \ddot{F}}{\partial \ddot{\mathbf{\Psi}}}\right\} &= \mathbb{E}_{\breve{\mathbf{H}}}\left\{P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\left[\frac{\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}}{(1+\tau)^2}+\breve{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}\right.\right. \\ &+ \left.\frac{\widehat{\mathbf{H}}^{\mathsf{H}}\breve{\mathbf{H}}+\breve{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}}{1+\tau}\right] + M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} \\ &- \left.\frac{P\varrho}{1+\tau}\widehat{\mathbf{H}} - P\varrho\breve{\mathbf{H}}\right\} = \mathbf{0} \\ &\overset{⑥}{\Longrightarrow} \ddot{\mathbf{\Psi}} = \widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \ddot{\varepsilon}\,\mathbf{I}\right)^{-1} \qquad (24)\end{aligned}$$

where ⑥ is due to equations (23a)–(23c), and therefore the regularization parameter $\ddot{\varepsilon}$ can now be expressed as

$$\begin{aligned}\ddot{\varepsilon} &= M(1+\tau)\left(\tau + \frac{\gamma^{-1}(1+\tau)^3\kappa}{\varrho}\right) \\ &= M(1+\tau) \\ &\times \left(\tau + \frac{\gamma^{-1}(1+\tau)^3(M+N)}{(N+1)+(M-1)\left(1+\sqrt{N\pi}+(M-2)\frac{\pi}{4}\right)}\right) \qquad (25)\end{aligned}$$

Remark 1: Since the Hessian matrix of the MSE objective function is positive definite, the expression in (24) is a global minimizer for the considered MMSE optimization problem. This implies the optimality of the derived regularization parameter in (25).

Remark 2: Note that by setting $\tau = 0$ as well as $M = N$, the adaptive RPA precoding boils down to the standard RPA precoding in [12]. In other words, adaptive RPA is a generalized version of the standard RPA without introducing any extra computational complexity.

V. Simulation Results

In this section and by using simulation results we certify the superior performance of the adaptive RPA precoding compared to CI and RCI precoding [4], PA precoding [6], and the standard RPA precoding [12]. For the case of digital feedback, the imperfect CSI, i.e., $\widehat{\mathbf{H}}$, is obtained by using RVQ as explained in [18]. Plus, in this case and as previously mentioned in Section III, we assume that the channel estimation error variance obeys $\tau = 2^{-B/N}$. In other words, when using adaptive RPA precoding under digital feedback, we replace τ by $2^{-B/N}$ in (25). For the case of reciprocal channels, we set $\tau = \beta\gamma^{-1}$, and the imperfect CSI, i.e., $\widehat{\mathbf{H}}$, is obtained by using (2). Also for the case of perfect CSI, adaptive RPA can be easily utilized by setting $\tau = 0$.

First in Fig. 2, we compare the performance of the adaptive RPA precoding with standard RPA precoding. Since the standard RPA precoding is only applicable to the case of equal number of transmit and receive antennas, we accordingly assume that $M = N = 8$. The BER results are depicted under QPSK signaling and for different CSI qualities including a reciprocal channel with $\beta = 10,\ \alpha = 1$, and CSI feedback scenarios based on RVQ using $B = 10,\ 20,$ and 30 fedback bits.

As seen under both digital and analog feedback, adaptive RPA precoding outperforms standard RPA. For example, when $\beta = 10,\ \alpha = 1$, adaptive RCI achieves nearly 6 dB gain in transmit power compared to standard RPA to reach the BER of 10^{-2}. Also under digital feedback, while the performance trend of standard RPA is nonmonotonic, that of the adaptive RPA is monotonic. As a result, at high SNRs, the achievable BER of adaptive RPA precoding based on $B = 10$ fedback bits is almost the same as the BER of standard RPA precoding based on $B = 30$ fedback bits.

This nonmonotonic behavior of standard RPA precoding is related to the case of digital feedback where at high SNRs, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

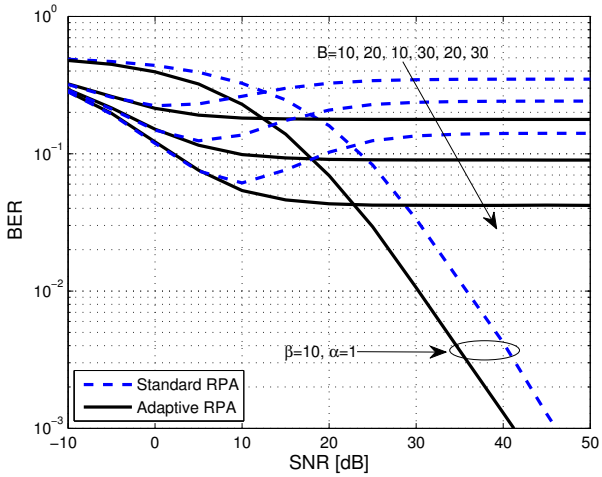

Fig. 2. BER of standard RPA and adaptive RPA precoding under various CSI qualities, for QPSK signaling, and for the case $M = N = 8$.

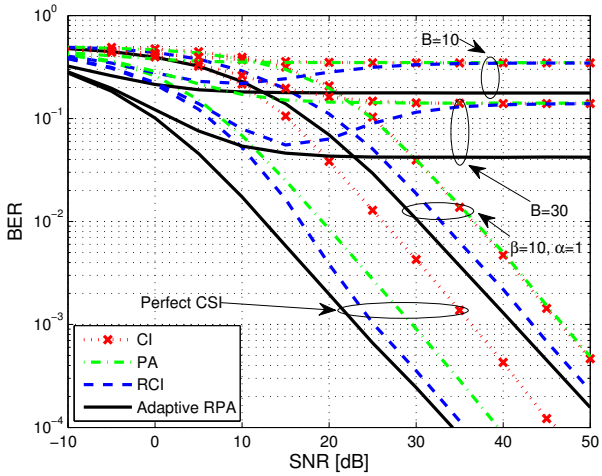

Fig. 3. Performance comparison of the BER of CI, RCI, PA, and adaptive RPA under various CSI qualities, for QPSK signaling, and for the case $M = N = 8$.

system becomes interference-limited. Therefore by increasing the SNR, the performance trend does not constantly improve and eventually becomes saturated. This implies that for standard RPA and by increasing the nominal SNR, the performance trend first becomes improved at low-to-intermediate SNRs, but suddenly deteriorates at a saddle point, and eventually becomes saturated at high SNRs, which leads to a nonmonotonic behavior. For the adaptive RPA precoding, on the other hand, the performance trend does not suddenly deteriorate. This is due to the regularization parameter $\ddot{\varepsilon}$ which is a function of the channel estimation error variance τ. Thus, in this case, the performance trend smoothly becomes saturated, which results in a monotonic behavior.

In Figs. 3–6, we compare the performance of the adaptive

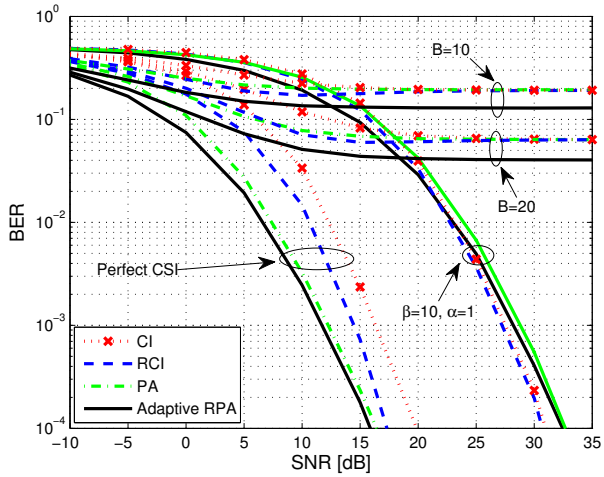

Fig. 4. Performance comparison of the BER of CI, RCI, PA, and adaptive RPA under various CSI qualities, for QPSK signaling, and for the case $M = 6,\ N = 8$.

RPA with CI, RCI, and PA precoding. We consider three different scenarios including the perfect CSI, the reciprocal channel ($\beta = 10,\ \alpha = 1$), and the CSI feedback.

Particularly in Fig. 3, we consider a scenario with QPSK signaling and for $M = N = 8$. For digital feedback, we assume that the number of fedback bits are $B = 10$ and $B = 30$. As seen, under both perfect and imperfect CSI, adaptive RPA outperforms other linear precoders. More specifically, under perfect CSI, adaptive RPA respectively achieves 4.5, 7, and 14 dB gain in transmit power compared to RCI, PA, and CI precoding to reach the BER of 10^{-2}. Meanwhile, under both reciprocal channels and CSI feedback, while PA precoding fails to outperform CI precoding especially at not-very-low SNRs, adaptive RPA precoding achieves a remarkable performance compared to RCI precoding. For example, under analog feedback, adaptive RPA achieves 2.5 dB gain in transmit power compared to RCI precoding to reach the BER of 10^{-2}. This gain is 6 dB compared to CI and PA precoding. Also under digital feedback, the BER curve of the RCI precoding demonstrates a nonmonotonic trend and as a result overlaps with those of CI and PA precoding at high SNRs; however, adaptive RPA manifests a distinguished performance. As demonstrated, at high SNRs, the BER of adaptive RPA based on $B = 10$ fedback bits is almost the same as the BER of CI, RCI, and PA precoding based on $B = 30$ fedback bits.

In Fig. 4, we consider a scenario with QPSK signaling and for $M = 6,\ N = 8$. For digital feedback, we assume that the number of fedback bits are $B = 10$ and $B = 20$. Note that in this case and unlike the depicted scenario in Fig. 3, the number of receive antennas is less than the number of transmit antennas. Nevertheless, under perfect CSI, adaptive RPA is still able to outperform CI, RCI, and PA precoding. This is also true for the case of digital feedback. However, in the case of reciprocal channels, while RPA achieves lower

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

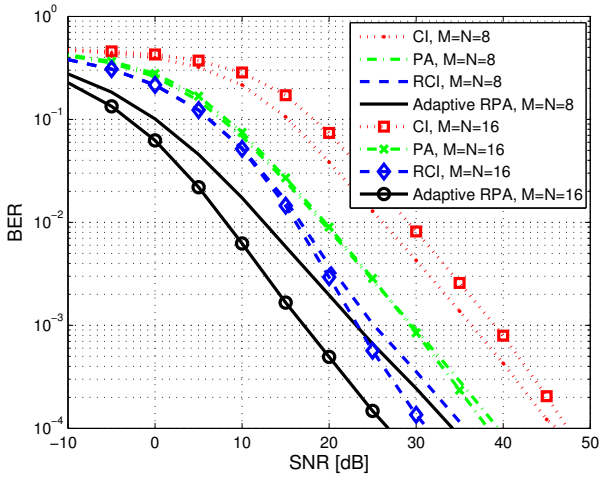

Fig. 5. Performance comparison of the BER of CI, RCI, PA, and adaptive RPA under perfect CSI, for QPSK signaling, and for the case $M = N = 8,\ M = N = 16$.

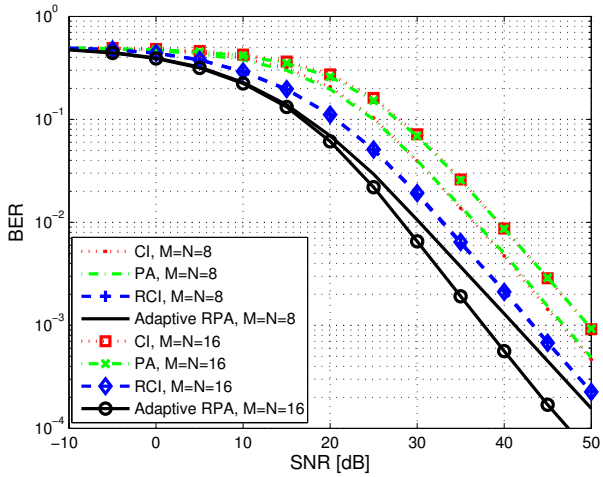

Fig. 6. Performance comparison of the BER of CI, RCI, PA, and adaptive RPA under a reciprocal channel with $\beta = 10,\ \alpha = 1$, for QPSK signaling, and for the case $M = N = 8,\ M = N = 16$.

BERs than CI, RCI, and PA at low-to-intermediate SNRs, its BER performance compared to CI and RCI precoders becomes slightly degraded at high SNRs.

Fig. 5 depicts the BER under QPSK signaling for CI, RCI, PA, and adaptive RPA precoding in the case of perfect CSI, and for $M = N = 8,\ M = N = 16$. With increasing M and N from 8 to 16, the following performance trends can be observed:

- For CI precoding, the BER deteriorates.
- For PA precoding, the BER remains almost constant.
- For RCI precoding, the BER remains constant at low-to-intermediate SNRs and improves slightly at high SNRs.
- For adaptive RPA precoding, the BER improves remarkably over all SNR ranges.

Fig. 6 depicts the BER under QPSK signaling for CI, RCI, PA, and adaptive RPA precoding in the case of a reciprocal channel with $\beta = 10,\ \alpha = 1$, and for $M = N = 8,\ M = N = 16$. With increasing M and N from 8 to 16, the following performance trends can be observed:

- For both CI and PA precoding, the BER deteriorates.
- For RCI precoding, the BER remains constant.
- For adaptive RPA precoding, the BER improves such that the larger the SNR, the lower BERs can be achieved.

This superior performance of RPA precoding is due to the fact that: First, by using phase alignment, RPA precoding increases the output SNR at each user for a fixed transmit power, compared to CI and RCI precoding. Second, since it inherently uses the idea of regularization, it improves the performance at low-to-intermediate SNRs, compared to CI and PA precoding.

Fig. 7 illustrates the BER of adaptive RPA precoding under various CSI qualities, for QPSK signaling, and for the case $M = 8,\ N = 8, 12, 16$. As seen, by keeping M fixed and increasing N, the BER performance remarkably improves.

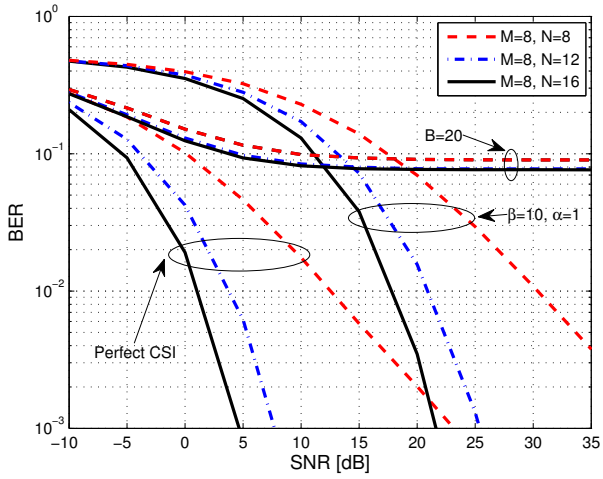

Fig. 7. Performance comparison of the BER of adaptive RPA precoding under various CSI qualities, for QPSK signaling, and for the case $M = 8,\ N = 8, 12, 16$.

This also highlights the fact that the case of $M = N$ achieves the worst BER.

VI. Conclusions

Thanks to multiple antennas, the BS is able to communicate to many users simultaneously at the expense of intra-cell interference. One effective way to suppress this interference is to deploy linear downlink precoding. However, to exploit full benefits of multiuser downlink, the availability of perfect CSI at the BS is necessary, which is not a realistic assumption in practice. In this case, performance of linear precoders may become severely degraded. Therefore, in this paper, we proposed an enhanced linear precoding scheme, dubbed adaptive RPA, by judiciously selecting an optimized regularization

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2372754, IEEE Transactions on Vehicular Technology

parameter based on the knowledge of the channel estimation error variance in advance. It was shown that the proposed RPA precoding is able to outperform standard linear precoders in the presence of both perfect and imperfect CSI, except the case of reciprocal channels with unequal number of transmit and receive antennas at high SNRs. Therefore, designing a modified version of adaptive RPA precoding which can achieve better performance for this very specific scenario might be of particular interest.

APPENDIX
STATISTICAL PROPERTIES OF $|\rho_{\ell,x}|$ (TABLE I)

Consider $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{M\times N}$ and $\mathrm{vec}\left(\boldsymbol{\mathcal{H}}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma_h^2\mathbf{I}\right)$. By expanding the complex multiplications of matrix $\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathsf{H}}$ for the case $\ell \neq x$, we have

$$\begin{aligned} |\rho_{\ell,x}| &= \left(\left[\sum_{n=1}^{N}\left(h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right)\right]^2 \right. \\ &+ \left.\left[\sum_{n=1}^{N}\left(h_{\ell,n}^i h_{x,n}^r - h_{\ell,n}^r h_{x,n}^i\right)\right]^2\right)^{\frac{1}{2}} \end{aligned} \quad (26)$$

where the notations $h_{\ell,n}^r = \Re\left(h_{\ell,n}\right), h_{\ell,n}^i = \Im\left(h_{\ell,n}\right)$, are used for convenience, and $h_{\ell,n}$ is used to denote the estimated channel coefficient of the n-th transmit antenna to the ℓ-th receive antenna, i.e., the (ℓ,n)-th element of $\boldsymbol{\mathcal{H}}$. Since we assumed that $h_{\ell,n}^r, h_{\ell,n}^i, h_{x,n}^r, h_{x,n}^i \in \mathcal{N}_{\mathbb{C}}\left(0, \frac{\sigma_h^2}{2}\right)$, we have

$$\mathbb{E}\left\{h_{\ell,n}^r h_{x,n}^r\right\} = 0 \quad \text{and} \quad \mathrm{var}\left\{h_{\ell,n}^r h_{x,n}^r\right\} = \frac{\sigma_h^4}{4} \quad (27)$$

The same applies to all combinations of real and imaginary coefficient that appear in (26). Therefore

$$\begin{aligned} \mathbb{E}\left\{h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right\} &= 0 \\ \mathrm{var}\left\{h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right\} &= \frac{\sigma_h^4}{2} \end{aligned} \quad (28)$$

and

$$\begin{aligned} \mathbb{E}\left\{\sum_{n=1}^{N}\left(h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right)\right\} &= 0 \\ \mathrm{var}\left\{\sum_{n=1}^{N}\left(h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right)\right\} &= \frac{N\sigma_h^4}{2} \triangleq \vartheta_1 \end{aligned} \quad (29)$$

Due to the symmetry of the real and imaginary parts of the channel taps, the values of (29) also apply to the second term on the right side of (26). Consequently $|\rho_{\ell,x}|$ is a Rayleigh variable with $\mathbb{E}\left\{|\rho_{\ell,x}|\right\} = \sqrt{2\vartheta_1}\Gamma\left(\frac{3}{2}\right) = \frac{\sigma_h^2\sqrt{N\pi}}{2}$ and $\mathbb{E}\left\{|\rho_{\ell,x}|^2\right\} = 2\vartheta_1\Gamma\left(2\right) = N\sigma_h^4$ [23], where $\Gamma\left(\cdot\right)$ is the gamma function such that $\Gamma\left(1\right) = 1$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, and $\Gamma\left(1+t\right) = t\Gamma\left(t\right)$.

For the case of $\ell = x$ we have

$$|\rho_{\ell,\ell}| = \sum_{n=1}^{N}|h_{\ell,n}|^2 = \sum_{n=1}^{N}\left[\left(h_{\ell,n}^r\right)^2 + \left(h_{\ell,n}^i\right)^2\right] \quad (30)$$

Since $h_{\ell,n}^r, h_{\ell,n}^i \in \mathcal{N}_{\mathbb{C}}\left(0, \frac{\sigma_h^2}{2}\right)$, $|\rho_{\ell,\ell}|$ is a χ-square random variable with $2N$ degrees of freedom, i.e., $|\rho_{\ell,\ell}| \sim \chi_{2N}^2$, and $\mathbb{E}\left\{|\rho_{\ell,\ell}|\right\} = 2N\vartheta_2 = N\sigma_h^2$ and $\mathrm{var}\left\{|\rho_{\ell,\ell}|\right\} = 4N\vartheta_2^2 = N\sigma_h^4$ [23] where $\vartheta_2 \triangleq \mathrm{var}\left\{h_{\ell,n}^r\right\} = \frac{\sigma_h^2}{2}$; therefore $\mathbb{E}\left\{|\rho_{\ell,\ell}|^2\right\} = \left[\mathbb{E}\left\{|\rho_{\ell,\ell}|\right\}\right]^2 + \mathrm{var}\left\{|\rho_{\ell,\ell}|\right\} = \sigma_h^4 N\left(N+1\right)$.

REFERENCES

[1] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[2] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[3] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Interference optimization for transmit power reduction in Tomlinson-Harashima precoded MIMO downlinks," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2470–2481, May 2012.

[4] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 192–202, Jan. 2005.

[5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[6] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252–262, Jan. 2011.

[7] F. A. Khan, C. Masouros, and T. Ratnarajah, "Interference-driven linear precoding in multiuser MISO downlink cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2531–2543, Jul. 2012.

[8] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.

[9] J. Zhang, C.-K. Wen, S. Jin, X. Gao, and K.-K. Wong, "Large system analysis of cooperative multi-cell downlink transmission via regularized channel inversion with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4801–4813, Oct. 2013.

[10] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[11] R. Muharar, R. Zakhour, and J. Evans, "Optimal power allocation and user loading for multiuser MISO channels with regularized channel inversion," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5030–5041, Dec. 2013.

[12] S. M. Razavi, T. Ratnarajah, and C. Masouros, "Transmit-power efficient linear precoding utilizing known interference for the multiantenna downlink," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4383–4394, Nov. 2014.

[13] V. K. Nguyen and J. S. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2008, pp. 1–4.

[14] S. M. Razavi and T. Ratnarajah, "Performance analysis of interference alignment under CSI mismatch," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4740–4748, Nov. 2014.

[15] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[16] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[17] ——, "Antenna combining for the MIMO downlink channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3834–3844, Oct. 2008.

[18] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3037–3046, Sep. 2012.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2006.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall, 1993.

[21] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[22] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[23] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2007.